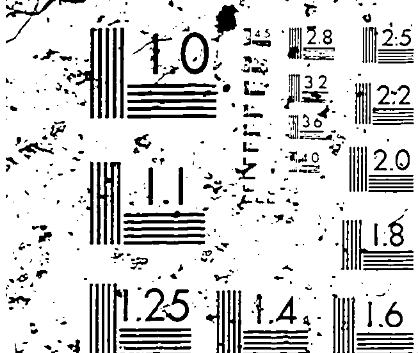

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS 1963-A

DOCUMENT RESUME

ED 191 855 TM 800 324

AUTHOR Koffler, Stephen L.; Penfield, Douglas A.
TITLE Nonparametric Discrimination Based Upon Inverse Normal Scores and Rank Transformations.
PUB DATE Apr 80
NOTE 34p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Discriminant Analysis; Hypothesis Testing; *Nonparametric Statistics
IDENTIFIERS *Inverse Normal Scores; *Rank Order Transformation

ABSTRACT

Two nonparametric statistical methods, the inverse normal scores method and the rank order transformation, are compared for use in discriminant function analysis. The methods are compared for both normal and non-normal distributions. When the distributions are normal, the rank and inverse normal scores methods are effective substitutes for the linear discriminant function (LDF) and the quadratic discriminant function (QDF). When the populations are non-normal, the LDF methods based on the ranks or the inverse normal scores are more effective than the LDF or QDF methods based on the raw data. Finally, when the criterion sample sizes are unequal, the inverse normal scores approach is more desirable than the rank approach. When the criterion sample sizes are equal, either of the two procedures can be used. (Author/JKS)

Nonparametric Discrimination Based Upon Inverse Normal Scores and Rank Transformations

Stephen L. Koffler
N.J. Department of Education

Douglas A. Penfield
Rutgers University

Presented at the 1980 annual meeting of the American Educational Research Association, Boston, Massachusetts

April, 1980

Nonparametric Discrimination Based Upon Inverse Normal Scores and Rank Transformations

Stephen L. Koffler
N.J. Department of Education

Douglas A. Penfield
Rutgers University

Behavioral science decisions frequently involve the rational assignment or classification of observations into one of a finite number of populations based on an evaluation of a series of measurements obtained on the observations. The set of statistical procedures that conventionally governs these decisions is known as discriminant analysis.

The theoretical basis for discriminant analysis was introduced by Welch (1939) who adapted the hypothesis testing concepts of Neyman and Pearson. Welch showed that a discrimination procedure that classified p-dimensional observations $\underset{\sim}{Z}$ into one of two populations Π_1 or Π_2 was equivalent to a partitioning of the sample space Ω into two mutually exclusive and exhaustive regions R_i $(i = 1,2)$, obtained by evaluating the likelihood ratio function at $\underset{\sim}{Z}$. $\underset{\sim}{Z}$ is assigned to Π_1 when the value of the likelihood ratio is greater than some appropriately determined constant κ, and to Π_2 when the value of the likelihood ratio is less than κ.[1]

It is possible that the classification decision for an observation could be in error; Z could originate from any population whose density is non-zero at Z. In the two population model, which will be the focus of this paper, two errors of classification are possible:

1. The procedure can assign $\underset{\sim}{Z}$ to Π_1 when $\underset{\sim}{Z}$ actually belongs to Π_2.
2. The procedure can assign $\underset{\sim}{Z}$ to Π_2 when $\underset{\sim}{Z}$ actually belongs to Π_1.

[1] When the likelihood ratio equals κ, the usual procedure is to randomly assign the observations to one of the populations.

Associated with each error is a probability of committing it (called the probability of misclassification), denoted by $P(\underset{\sim}{Z} \in \Pi_i \mid \Pi_j)$, $(i \neq j = 1,2)$.

Welch(1939) showed that for the two population situation, with observations drawn from known distributions, the optimal solution to the classification problem is

$$f_1(\underset{\sim}{Z}) \,/\, f_2(\underset{\sim}{Z}) \qquad (1)$$

where $f_i(\underset{\sim}{Z})$ is the density function of the ith distribution evaluated at $\underset{\sim}{Z}$. $\underset{\sim}{Z}$ is classified into Π_1 if (1) is greater than constant κ, and into Π_2 if (1) is less than κ. Equation (1) is optimal in the sense that it minimizes $P(\underset{\sim}{Z} \in \Pi_i \mid \Pi_j)$. κ is defined as

$$\kappa = C_{12}q_2 \,/\, C_{21}q_1 \qquad (2)$$

where C_{ij} $(i \neq j = 1,2)$ is the cost of misclassifying an observation from Π_j into Π_i and q_i $(i = 1,2)$ is the a priori probability that the observation belongs to population Π_i (Anderson, 1951).

Procedures for Multivariate Normal Distributions

Frequently data are collected that are representative of multivariate normal distributions. When the populations are so distributed with known mean vectors and identical covariance matrices, (1) simplifies to

$$\underset{\sim}{Z}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2) - \tfrac{1}{2}(\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2)'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2) \qquad (3)$$

where $\underset{\sim}{\Sigma}$ is the common covariance matrix and $\underset{\sim}{\mu}_i$ $(i = 1,2)$ is the mean vector of Π_i. When (3) is greater than $\log(\kappa)$, $\underset{\sim}{Z}$ is classified as belonging to Π_1;

when (3) is less than log (κ), $\underset{\sim}{Z}$ is classified as belonging to Π_2.[2] Equation (3) is referred to as the Linear Discriminant Function(LDF).

If the data originate from multivariate normal distributions with known parameters, but the covariance matrices are not identical for the two populations, the form of the likelihood ratio is quadratic.

$$\tfrac{1}{2}\underset{\sim}{Z}'(\underset{\sim}{\Sigma}_2^{-1} - \underset{\sim}{\Sigma}_1^{-1})\underset{\sim}{Z} + (\underset{\sim}{\mu}_1'\underset{\sim}{\Sigma}_1^{-1} - \underset{\sim}{\mu}_2'\underset{\sim}{\Sigma}_2^{-1})'\underset{\sim}{Z} +$$
$$\tfrac{1}{2}([\underset{\sim}{\mu}_2'\underset{\sim}{\Sigma}_2^{-1}\underset{\sim}{\mu}_2 - \underset{\sim}{\mu}_1'\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\mu}_1] - \log[|\underset{\sim}{\Sigma}_1| / |\underset{\sim}{\Sigma}_2|]) \qquad (4)$$

where $\underset{\sim}{\mu}_i$ and $\underset{\sim}{\Sigma}_i$ $(i = 1,2)$ are the mean vector and covariance matrix, respectively, from Π_i. Equation (4) is called the Quadratic Discriminant Function(QDF) and its classification decision is identical to that of the LDF. Marks & Dunn(1974) have shown that under the assumption of multivariate normality and unequal covariance matrices, the QDF misclassifies fewer observations than the LDF.

Equation (1) is optimal only when the densities are known and completely specified. It is infrequent, however, that researchers encounter situations where the distributions from which their data are drawn are completely specified. Usually, the densities are either completely unknown or are known except for one or more parameters. For these situations, the unknown parameters must be estimated from samples and procedures based on the sample estimates developed to classify new observations.

Hoel & Peterson(1949) and Fix & Hodges(1951) determined that the best sample based procedure is of the likelihood-ratio type where the sample estimates replace the unknown parameters (called "plug-in" procedures).

---

[2] The classification constant is log (κ) because of algebraic simplification. When $C_{12} = C_{21}$ and $q_1 = q_2$, log (κ) = 0.

Anderson (1951) developed a statistic that is the sample based analogue to the LDF for data sampled from multivariate normal distributions with a common but unknown covariance matrix and unknown mean vectors. Anderson's statistic is in the form of (3) with the maximum likelihood estimates ($\bar{X}_i$ and S) substituted for the mean vector and identical covariance matrix, respectively. Using a similar argument, a sample based Quadratic Discriminant Function is analogous to (4), with the maximum likelihood estimates ($\bar{X}_i$ and S_i) replacing the unknown mean vectors and unequal covariance matrices.

Procedures for Unknown or Non-Normal Distributions

The LDF and QDF are optimal only when the data are multivariate normal. In the past, however, researchers have relied upon the sample based LDF or QDF to resolve the classification problem, regardless of the underlying distributions of the data. Lachenbruch, Sneeringer & Revo(1973), Johnson & Ramberg (1978) and Koffler & Penfield(1979) have investigated the robustness of the LDF and QDF when classifying observations from non-normal distributions The three studies showed that when data were drawn from such distributions, the proportion of observations misclassified using the LDF or QDF was substantially altered from what was expected. Thus, the LDF and QDF are not robust to the normality assumption and researchers could be misled by using either procedure when investigating non-normal distributions.

Several nonparametric procedures have been suggested as possible alternatives to the LDF or QDF for non-normal data. Koffler & Penfield(1979) have empirically compared several nonparametric procedures. That study showed that procedures such as the Nearest Neighbor with Probability Blocks (Anderson, 1966; Fix & Hodges, 1951; Gessaman & Gessaman, 1972) and the

Loftsgaarden-Quesenberry density estimator (Loftsgaarden & Quesenberry, 1972) classified observations equally as effectively as either the LDF or QDF when data were sampled from multivariate normal distributions and better than either procedure when data were sampled from non-normal distributions. However, these nonparametric procedures have limited utility because they generally require larger samples from each population since the entire density function must be estimated rather than simply unknown parameters.

Conover & Iman (1978) have suggested another solution to the classification problem for non-normal data based on first transforming the data to make the distribution functions approximately normal and then applying the sample based LDF or QDF to the transformed data. This procedure is considerably simpler to use than the other nonparametric alternatives and requires only one step more than the LDF or QDF, namely, the ranking of the data and observations to be classified.

Conover & Iman empirically constrasted their suggested classification procedures with other procedures (including the nonparametric ones of Koffler & Penfield and the LDF and QDF using the original data). They concluded that if the data were normally distributed, the rank methods performed equally as well as the LDF and QDF; if the data were non-normal, the rank transformation method worked better than the LDF or QDF and as well as any of the nonparametric alternatives. In all instances, the measure of performance considered was the overall proportion of misclassified observations.

It should not be surprising that the rank transformation was an appropriate one for the data. Many nonparametric procedures, such as the Mann-Whitney (Wilcoxon) test, the Kruskal-Wallis test and the Spearman rank, order correlation, are based upon rank transformations and have been shown to be effective alternatives to their parametric counterparts. Furthermore, the rank transformation has been shown to work effectively in multivariate regression analysis (Iman & Conover, 1977) and in the analysis of experimental data (Iman, 1974; Conover & Iman, 1976).

Normal Scores Type Transformations

A natural extension of the Conover & Iman (1978) study involves the investigation of alternative transformations that could be used to effectively classify data from all types of distributions.

The normal scores transformation is one that should be considered. This type of transformation derives its values from various properties of the normal distribution. Two forms of the transformation are usually considered: the expected normal order statistic (Hoeffding, 1951; Terry, 1952) and the inverse normal score (Van der Waerden, 1952, 1953, 1956).

Tests based on normal scores transformations have not been used as frequently as those based upon ranks. However, the results from those instances where such transformations have been applied suggest that they have utility in a number of situations, specifically for discriminant analysis.

The efficiency of one test (T_1) relative to another (T_2) can be determined by comparing the ratio of n_2/n_1, where n_i is the sample size of T_i, ($i = 1,2$), under the condition that both tests are used to test a specific hypothesis, have identical α and β levels and, therefore, are comparable with respect to level of significance and power (Conover, 1971).

Tests based upon the normal scores transformation have been extensively used for the k-sample location problem ($k \geq 2$). For the two sample problem the Mann-Whitney (Wilcoxon) statistic has an efficiency relative to the t-test of 95.5% for normal distributions, 100% for uniform distributions, and may be infinite for other distributions.

A similar test which utilizes a normal scores transformation has an asymptotic relative efficiency to the t-test of 100% when the t-test assumptions are satisfied, and greater than 100% when the t-test assumptions are violated. The normal scores test is more efficient than the Mann-Whitney (Wilcoxon) test when the distributions break off abruptly (e.g. uniform or exponential), the rank test is more efficient for distributions with heavy tails (e.g. logistic or Cauchy), and there is essentially no difference between the two tests when the distributions are approximately normal (Lehmann, 1975).

When $k \geq 3$, the Kruskal-Wallis test is generally used when the assumptions of the one-way analysis of variance F test are not satisfied. Hájek & Sidak(1967) derived test statistics based upon expected normal order statistics and inverse normal scores. Puri(1964) showed that the asymptotic relative efficiency of Hajek & Sidak's normal scores test relative to the Kruskal-Wallis test or to the F test is the same as that of the two sample normal scores test relative to the Mann-Whitney (Wilcoxon) test or t-test. Furthermore, Pratt(1964) has shown that these normal scores tests are far less sensitive to non-homogeneity of variance than is the F test or the Kruskal-Wallis test.

Because of the efficacy of the normal scores transformation for the location problem and its superiority to the rank transformation in certain situations, it is of value to determine whether procedures based on these transformations can be used to resolve the classification problem for data sampled from non-normal distributions. A natural extension to the Conover & Iman(1978) study is an investigation of the effectiveness of classifying observations with the LDF and QDF based upon a normal scores transformation.

The purpose of the research described in this paper is to empirically contrast classification procedures based on normal scores with those based upon ranks and upon the original data when the data originate from both normal and non-normal distributions.

Methodology

To estimate the LDF and QDF parameters, criterion samples of varying sizes were generated for four types of two dimensional distributions. The four distributions considered were the bivariate normal distribution and non-normal representatives from three classes of distributions: 1) finite range (logit normal); 2) semi-infinite range (log normal); and 3) infinite range (inverse hyperbolic sine normal). In all instances the two dimensions were independent.

The three non-normal distributions were generated from the Johnson(1949) system of distributions. To obtain the required non-normal samples, normally distributed random variables were generated and then the appropriate inverse transformation applied. The Johnson system of transformations is summarized in Table 1. In Table 1 the variable x is normally distributed, while the variable y is distributed according to the appropriate non-normal distribution. An algorithm by Ramberg & Schmeiser (1972), based upon the inverse function of the lambda distribution, was used to generate the normal deviates. Random deviates from a uniform distribution were needed to obtain the normal deviates. A multiplicative congruential procedure developed by Kossack & Henschke (1975) was used for this purpose.

TABLE 1

TRANSFORMATIONS (AND THEIR INVERSES) THAT GENERATE THE JOHNSON (1949) SYSTEM OF DISTRIBUTIONS

DISTRIBUTION	TRANSFORMATION		INVERSE
Log Normal	$y = \log x$	$0 \leq x \leq \infty$	$x = \mathrm{EXP}(y)$
Logit Normal	$y = \log(x/1-x)$	$0<x<1$	$x = \mathrm{EXP}(y)/1 + \mathrm{EXP}(y)$
Inverse Hyperbolic Sine Normal	$y = \mathrm{Sinh}^{-1}(x)$	$-\infty<x<\infty$	$x = \mathrm{Sinh}(y)$

The bivariate normal distributions that were used to generate the non-normal samples for Π_1 and Π_2 each had the identity matrix for its covariance matrix. The mean vector for Π_1 was $(\mu,0)$ and for Π_2 it was $(0,0)$. For each of the four distributions, samples were generated for each combination of sample size [(n_1,n_2) = (8,8), (8,27), (8,64), (8,200), (27,27), (27,64), (27,200), (64,64), (64,200), (200,200)] and first component of the mean vector for Π_1 (μ = 1,2).[3] In total there were samples drawn from twenty combinations of (n_1,n_2) and μ for each distribution.

The sample based LDF and QDF were used to establish the classification rules, assuming equal costs of misclassification and equal a priori

[3] These values of n_1,n_2, and μ were selected to parallel previous studies, including those of Conover & Iman (1978) and Koffler & Penfield (1979).

probabilities of group membership (i.e. log (K) = 0). As previously outlined, both the LDF and QDF involve the estimation of the population means and the covariance matrix (either pooled for the LDF or separate for the QDF) from the criterion samples. These estimated values are then substituted into (3) and (4). The parameter estimates for the LDF and QDF were obtained in three ways, using the raw data, the ranks of the data and the corresponding inverse normal scores.[4]

Once the LDF and QDF parameters were estimated for each combination of sample size and first component of the mean vector for Π_1, index samples consisting of 1000 new observations from each original population were generated. For each data point, the rank and the inverse normal score were computed. Each value of the index samples was entered into (3) and (4) and the classification of the value determined. The proportion of misclassified observations for each sample and over all samples was obtained. In all there were six classification methods studied (the LDF and QDF based on the raw data, ranks, and inverse normal scores).

The process was repeated 20 times. Thus, the population parameters were estimated 20 different times and each time 2000 observations were classified. The estimated probability of misclassification for each sample was based on 20,000 observations and the overall estimated probabilities of misclassification were based on 40,000 classifications for each combination of n_1, n_2, and μ.[5]

---

[4] The inverse normal scores transformation differs little from the expected normal order statistic transformation. The two transformations are asymptotically equivalent and structurally identical (McSweeney & Penfield, 1968). The inverse normal scores transformation was used because of its ease of computation.

[5] All computer programs to generate the data and classification procedures were written in the FORTRAN IV programming language.

To obtain the ranks of the data, the two criterion samples of size n_1 and n_2 were combined. All observations in each of the two dimensions were then replaced by their corresponding rank; rank 1 for the smallest observation to rank N ($N = n_1 + n_2$) for the largest observation in each dimension. Each dimension was ranked separately and ranks of tied observations were assigned randomly.

To obtain the rank for each of the 1000 observations in the index samples, each new observation was compared dimension by dimension with all N original observations. For each dimension of the new observation, the original score was replaced by a number obtained by linear interpolation between two adjacent ranks from the original criterion samples. These interpolated ranks represented the placement of that dimension among the corresponding values of the same dimension in the N criterion sample observations. (Conover & Iman, 1978).

The derivation of the Van der Waerden inverse normal scores transformation is based upon the ranks of the data. For this transformation, assume the rank of the ith largest observation in a particular dimension is denoted by R_i and $\Phi(X)$ represents the cumulative distribution function of a standard normal random variable. The Van der Waerden transformation is derived first by dividing each of the ranks R_i by the quantity $(N + 1)$. This creates a distribution of scores in the interval $(0,1)$. Then, by considering $R_i/(N + 1)$ as a percentile of a normal distribution (i.e. $\Phi(X_i) = R_i/(N + 1)$), the X_i values can be determined by performing the inverse operation. That is, if $\Phi(X_i) = R_i/(N + 1)$, then $X_i = \Phi^{-1}(R_i/(N + 1))$. The X_is form the Van der Waerden inverse normal scores.

Optimal Probability of Misclassification

For each of the distributions, it is possible to determine the optimal probability of misclassification (i.e. $P(\underset{\sim}{Z} \in \Pi_i \mid \Pi_j)$, when the population parameters are completely specified and the distributions known).[6] Anderson (1951) has shown that the optimal probability of misclassification associated with the LDF is denoted by $\Phi(-\Delta/2)$, where Δ^2 is the Mahalanobis distance between the two populations.[7] Since the multivariate normal distributions in the present study were independent and the only non-zero mean component is μ, Δ^2 simplifies to μ^2. Hence, $\Phi(-\Delta/2) = \Phi(-\mu/2)$.

The corresponding values for the optimal probability of misclassification for the bivariate normal distributions under study are $\Phi(-1/2) = 0.3085$ and $\Phi(-2/2) = 0.1587$. Anderson(1951) has shown that the LDF minimizes the sum of the individual probabilities of misclassification (i.e. $P(1/2)+P(2/1)$). In the case of multivariate normal distributions, this occurs when $P(1/2) = P(2/1)$.

Since the non-normal data were obtained from non-linear transformations of data drawn from bivariate normal distributions, those data can be transformed back to the bivariate normal distributions by performing the inverse operation. The optimal classification procedure for the non-normal data involves transforming the data to normality and then applying the LDF. Thus, the optimal probability of misclassification for the non-normal data is identical to that of the original bivariate normal distributions.

---

[6] For simplicity, let $P(\underset{\sim}{Z} \in \Pi_1 \mid \Pi_2) = P(1/2)$, $P(\underset{\sim}{Z} \in \Pi_2 \mid \Pi_1) = P(2/1)$ and $P =$ the overall error rate.

[7] This is true when there are equal costs of misclassification, equal a priori probabilities of group membership and completely specified multivariate normal distributions with equal covariance matrices.

For each combination of n_1, n_2, and μ, the proportion of misclassified observations, $\hat{P}(1/2)$, $\hat{P}(2/1)$, and $\hat{P}$, were determined and served as the performance criteria and means of comparison among the six procedures.[8] Given the optimal values for the probabilities of misclassification, the effectiveness of the six sample based procedures can be determined by comparing the proportions of misclassification to the optimal rate. The empirically determined proportions of misclassification are estimates of the optimal values, and the procedure that provides the best estimates is considered to be most effective.

Two criteria for comparison were considered. The first was the relative disparity between $\hat{P}(1/2)$ and $\hat{P}(2/1)$ for each of the procedures. The smaller the disparity, the more effective the procedure (assuming that the overall proportion of misclassified observations approached the optimal probability). The second criterion was the overall error rate. These values were compared with the optimal values and a close agreement indicated an effective procedure. Tests of proportion and associated post hoc procedures (Marascuilo, 1966) were used to analyze the data. It is important to note that in many instances while the differences among the proportions were small, they were statistically significant ($p < .05$).

Results

The results for each of the four distributions are presented in Tables 2,4,5, and 6.[9] An examination of the tables shows that the

---

[8] $\hat{P}(i/j)$ is the empirically determined estimate of $P(i/j)$. $\hat{P}$ is the overall error rate and is equal to $[\hat{P}(1/2) + \hat{P}(2/1)]/2$ because equal numbers of observations were classified from each sample. These estimates represent the average proportion of misclassified observations for the 20 trials.

[9] The following abbreviations are used for the remainder of the paper: LDF = LDF procedure based on the raw data; RLDF = LDF procedure based on the ranks; ILDF = LDF procedure based on the inverse normal scores. A similar set of abbreviations are used for the QDF procedures.

proportion of misclassified observations for the rank transformation or for the inverse normal scores transformations was identical for all of the distributions. This is to be expected since the non-normal data were derived from monotonic transformations of the normal data. Because of the monotonicity of the transformations, the order of the data remained unchanged regardless of the distribution. Therefore, the ranks and inverse normal scores of the original data were unchanged, the sample estimates of the population parameters were likewise unchanged, and the classification decisions were identical.

Normal Distribution

Table 2 presents the results for the bivariate normal samples of data. For these data, it was expected that the performance of the LDF and QDF should be almost identical because the covariance matrices were estimated from populations both having the identity covariance matrix. From (4) it is evident that when the two covariance matrices are identical, the QDF is equivalent to the LDF.

$\mu = 1$

When $n_1 = 8$, the three procedures based on the LDF had approximately the same overall proportion of misclassified observations and discrepency between $\hat{P}(1/2)$ and $\hat{P}(2/1)$. In all cases, the estimated overall proportion of misclassification was significantly greater than the optimal value of 0.3085. This result was not unexpected since the sample mean and covariance estimates for Π_1 were based upon eight observations and thus had a large standard error.

The three procedures based on the QDF misclassified considerably more

TABLE 2

RESULTS FOR THE NORMAL DISTRIBUTION
TOTAL PROPORTION OF MISCLASSIFIED OBSERVATIONS

(N1,N2)		(8,8)			(8,27)			(8,64)			(8,200)			(27,27)		
MEAN	PROCEDURE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
1	LDF	0.397	0.314	0.355	0.343	0.310	0.327	0.361	0.285	0.323	0.344	0.311	0.327	0.328	0.307	0.318
	QDF	0.457	0.356	0.406	0.390	0.329	0.360	0.425	0.289	0.357	0.435	0.288	0.361	0.337	0.315	0.326
	RLDF	0.382	0.328	0.355	0.321	0.340	0.330	0.325	0.322	0.324	0.301	0.359	0.330	0.335	0.298	0.316
	RQDF	0.450	0.341	0.396	0.348	0.350	0.349	0.343	0.336	0.339	0.350	0.350	0.350	0.345	0.296	0.320
	ILDF	0.376	0.329	0.353	0.339	0.323	0.331	0.342	0.305	0.323	0.330	0.371	0.326	0.335	0.301	0.318
	IQDF	0.457	0.334	0.396	0.372	0.338	0.355	0.398	0.308	0.353	0.407	0.304	0.356	0.342	0.302	0.322

(N1,N2)		(27,64)			(27,200)			(64,64)			(64,200)			(200,200)		
MEAN	PROCEDURE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
1	LDF	0.318	0.308	0.313	0.334	0.285	0.309	0.313	0.309	0.311	0.299	0.313	0.306	0.310	0.315	0.313
	QDF	0.335	0.324	0.329	0.359	0.285	0.322	0.294	0.337	0.316	0.313	0.306	0.309	0.315	0.312	0.314
	RLDF	0.303	0.321	0.312	0.282	0.336	0.309	0.317	0.301	0.309	0.265	0.348	0.307	0.301	0.322	0.312
	RQDF	0.311	0.324	0.318	0.275	0.354	0.315	0.309	0.317	0.313	0.255	0.359	0.307	0.305	0.320	0.312
	ILDF	0.321	0.307	0.314	0.324	0.295	0.310	0.318	0.304	0.311	0.287	0.325	0.306	0.303	0.321	0.312
	IQDF	0.335	0.318	0.327	0.337	0.304	0.321	0.304	0.326	0.315	0.293	0.323	0.308	0.307	0.319	0.313

(N1,N2)		(8,8)			(8,27)			(8,64)			(8,200)			(27,27)		
MEAN	PROCEDURE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
2	LDF	0.203	0.172	0.188	0.168	0.171	0.170	0.177	0.159	0.163	0.168	0.161	0.165	0.170	0.161	0.166
	QDF	0.208	0.196	0.202	0.204	0.172	0.188	0.232	0.154	0.193	0.233	0.151	0.192	0.170	0.167	0.168
	RLDF	0.202	0.178	0.190	0.118	0.231	0.174	0.105	0.253	0.179	0.088	0.279	0.184	0.157	0.172	0.164
	RQDF	0.224	0.178	0.201	0.150	0.207	0.178	0.178	0.199	0.189	0.180	0.207	0.193	0.157	0.172	0.165
	ILDF	0.212	0.184	0.198	0.146	0.198	0.172	0.143	0.199	0.171	0.137	0.194	0.165	0.157	0.175	0.166
	IQDF	0.235	0.184	0.209	0.183	0.191	0.187	0.207	0.175	0.191	0.220	0.170	0.195	0.160	0.177	0.169

(N1,N2)		(27,64)			(27,200)			(64,64)			(64,200)			(200,200)		
MEAN	PROCEDURE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
2	LDF	0.181	0.154	0.167	0.167	0.156	0.161	0.163	0.156	0.160	0.161	0.161	0.161	0.158	0.157	0.159
	QDF	0.190	0.151	0.170	0.172	0.156	0.164	0.162	0.159	0.161	0.161	0.159	0.160	0.160	0.157	0.159
	RLDF	0.130	0.209	0.162	[illegible]	0.271	0.170	0.162	0.159	0.160	0.102	0.234	0.168	0.165	0.153	0.159
	RQDF	0.139	0.200	0.170	0.106	0.236	0.171	0.162	0.160	0.161	0.108	0.225	0.167	0.165	0.153	0.159
	ILDF	0.156	0.185	0.171	0.125	0.205	0.166	0.163	0.160	0.162	0.129	0.197	0.163	0.164	0.154	0.159
	IQDF	0.162	0.181	0.172	0.136	0.193	0.165	0.164	0.161	0.163	0.131	0.198	0.164	0.165	0.155	0.160

observations than those based on the LDF. The fact that the covariance matrix for Π_1 was based on so few observations also provides an explanation for the large differences between the LDF and QDF type procedures - the pooled sample covariance matrix was probably very different from the separate covariance matrices.

When $n_1 = n_2 > 8$, all six procedures misclassified approximately the same proportion of observations and were approximately equivalent to the optimal value. In all cases, the difference between $\hat{P}(1/2)$ and $\hat{P}(2/1)$ was smallest for the LDF based on the original data; however, the discrepency for all of the procedures was similar.

When $(n_1, n_2) = (27,64)$ or $(27,200)$, the three LDF procedures minimized the proportion of misclassified observations; however, the proportion of overall errors for the three QDF procedures did not differ substantially from the ones for the LDF, especially for the RQDF. When $n_1 = 27$ and $n_2 = 64$, the difference between $\hat{P}(1/2)$ and $\hat{P}(2/1)$ was approximately equal for all of the procedures; when $n_1 = 27$ and $n_2 = 200$, the discrepencies were smallest for the ILDF and IQDF. However, none of the procedures exhibited discrepencies that were extensive. When $(n_1, n_2) = (64,200)$, all of the procedures were equally as effective in terms of the overall error rate. For the relative discrepency between $\hat{P}(1/2)$ and $\hat{P}(2/1)$, the LDF and QDF, based on the original data, minimized the difference, while the RLDF and RQDF exhibited a relatively severe inflation/deflation phenomenon (i.e. $\hat{P}(1/2)$ was considerably smaller than the optimal value while $\hat{P}(2/1)$ was considerably larger).

$\mu = 2$

There was less of a disparity among the six procedures when $\mu = 2$

than when $\mu = 1$ for the bivariate normal data. When $n_1 = 8$, the three procedures based on the LDF were again most effective in minimizing the overall error rate. However, for the two largest values of n_2, the RLDF procedure was not as accurate as the LDF or ILDF. Additionally, the discrepency between $\hat{P}(1/2)$ and $\hat{P}(2/1)$ was considerably larger for the RLDF than for the LDF or ILDF.

When $n_1 = n_2 > 8$, there was no appreciable difference among the overall error rates or the discrepencies between $\hat{P}(1/2)$ and $\hat{P}(2/1)$ for the six procedures. Additionally, as the sample size increased, $\hat{P}$ approached 0.1587, the optimal error rate. For the situation when $n_1 \neq n_2$, there was no discernible difference among the overall error rates when $(n_1, n_2) = (27,64)$ or $(64,200)$; however, when $(n_1, n_2) = (27,200)$, the RLDF and RQDF classified larger numbers of observations incorrectly than the other four procedures. Additionally, for all three of these sample sizes, the RLDF and RQDF discrepencies were significantly larger than the discrepency for the other four procedures.

Summary

As expected, the LDF based on the original data proved to be an effective classification procedure for the bivariate normal data. Furthermore, as the sample sizes increased, the QDF effectively classified the data because the separate covariance estimates and pooled covariance estimates began to converge to the identity matrix.

In all instances, the RLDF and ILDF proved to be as effective as the LDF. The only exception to this occurred for the RLDF when the sample sizes were most disparate. When the sample sizes were equal, there was no discernible difference in the classification ability of the three LDF methods.

Non-Normal Distributions

Table 3 illustrates the means and variances for each dimension of the three non-normal distributions. Clearly, the variances for Π_1 are markedly different from that for Π_2 for each of the non-normal samples. The difference between the two populations is only in the first dimension of the mean vector, however, this affects the entire classification process through the sample covariance matrix. It was therefore appropriate to consider classification according to the QDF procedure for these data.

Recall that the procedures based upon the ranks and the inverse normal scores were identical for all of the non-normal distributions and for the bivariate normal distributions because the transformations were monotonic. Because of that, the three non-normal distributions can be considered together with respect to the classification of the index data based on the rank and inverse normal scores procedures. They must, however, be considered separately with respect to the LDF and QDF based on the original data.

An examination of the results reveals that the LDF and QDF classified both the log normal and inverse hyperbolic sine normal data similarly, while they classified the logit normal data differently from the other two, but similarly to the bivariate normal data. Hence, the results for the classification of the log normal and inverse hyperbolic sine normal data will be discussed together and the logit normal data separately.

TABLE 3

MEANS AND VARIANCES OF THE NON-NORMAL DISTRIBUTIONS
FOR SPECIFIED MEANS OF THE NORMAL DISTRIBUTION
$(\sigma_y^2 = 1)$[a]

	Log Normal	
μ_y	η_x	σ_x^2
0	1.65	4.67
1	4.48	34.51
2	12.18	255.02
	Logit Normal	
μ_y	η_x	σ_x^2
0	0.50	0.043
1	0.70	0.029
2	0.84	0.019
	Inverse Hyperbolic Sine Normal	
μ_y	η_x	σ_x^2
0	0.00	3.19
1	1.94	9.65
2	5.98	74.63

SOURCE: Lachenbruch, Sneeringer and Revo. Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-Normality. _Communications in Statistics_, 1973, _1_, 54.

[a] σ_y^2 is the variance of the underlying normal distribution.

μ_y is the mean of the underlying normal distribution.

η_x is the mean of the transformed non-normal variate.

σ_x^2 is the variance of the transformed non-normal variate.

Log Normal & Inverse Hyperbolic Sine Normal Distributions

The results for these data appear in Tables 4 and 5. For all combinations of sample size and μ, the LDF and QDF based on the original data significantly misclassified more observations than the procedures based upon the rank or inverse normal scores transformations. The LDF and QDF based on the original data further exhibited a severe inflation/deflation effect. The LDF and QDF were clearly inappropriate for these types of non-normal distributions. Thus, the remaining discussion will consider only the four nonparametric procedures.

$\mu = 1$

For $n_1 = 8$, the RLDF and ILDF procedures minimized the overall error rate; for all other combinations of sample size, all of the four nonparametric procedures were equally as effective, with the exception of the IQDF when $(n_1, n_2) = (27,64)$ or $(27,200)$. The inflation/deflation effect related to the discrepency between $\hat{P}(1/2)$ and $\hat{P}(2/1)$ was smallest for the ILDF; however, in most instances, there was little difference among the four procedures.

$\mu = 2$

The pattern for this value of μ was essentially identical to the pattern when $\mu = 1$. When $n_1 = 8$, the RLDF and ILDF were most effective in minimizing the overall error rate. As the sample size increased, the four procedures became indistinguishable in terms of $\hat{P}$ and $\hat{P}$ approached the optimal rate of 0.1587. However, upon examination of $\hat{P}(1/2)$ and $\hat{P}(2/1)$, it became apparent that the inflation/deflation effect was substantial for both the RLDF and RQDF in many instances. The discrepency for

TABLE 4.

RESULTS FOR THE LOG NORMAL DISTRIBUTION
TOTAL PROPORTION OF MISCLASSIFIED OBSERVATIONS

,N2)	(8 ,8)			(8 ,27)			(8 ,64)			(8 ,200)			(27 ,27)		
JRE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
DF	0.543	0.248	0.395	0.510	0.207	0.359	0.596	0.158	0.377	0.547	0.151	0.349	0.542	0.160	0.351
DF	0.544	0.286	0.415	0.499	0.290	0.395	0.579	0.293	0.436	0.587	0.212	0.399	0.578	0.198	0.388
DF	0.382	0.328	0.355	0.321	0.340	0.330	0.325	0.322	0.324	0.301	0.359	0.330	0.335	0.298	0.316
DF	0.450	0.341	0.396	0.348	0.350	0.349	0.343	0.336	0.339	0.350	0.350	0.350	0.345	0.296	0.320
DF	0.376	0.329	0.353	0.339	0.323	0.331	0.342	0.305	0.323	0.330	0.321	0.326	0.335	0.301	0.318
DF	0.457	0.334	0.396	0.372	0.333	0.355	0.398	0.308	0.353	0.407	0.304	0.356	0.342	0.302	0.322

,N2)	(27 ,64)			(27 ,200)			(64 ,64)			(64 ,200)			(200 ,200)		
JRE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
DF	0.545	0.150	0.348	0.574	0.122	0.348	0.514	0.179	0.347	0.535	0.143	0.339	0.545	0.138	0.342
DF	0.613	0.152	0.383	0.633	0.124	0.378	0.648	0.104	0.376	0.622	0.120	0.371	0.664	0.088	0.376
DF	0.303	0.321	0.312	0.282	0.336	0.307	0.317	0.301	0.309	0.265	0.348	0.307	0.301	0.322	0.312
DF	0.311	0.324	0.318	0.275	0.354	0.315	0.309	0.317	0.313	0.255	0.359	0.307	0.305	0.320	0.312
DF	0.321	0.307	0.314	0.324	0.296	0.310	0.318	0.304	0.311	0.287	0.325	0.306	0.303	0.321	0.312
DF	0.335	0.318	0.327	0.337	0.304	0.321	0.304	0.326	0.315	0.293	0.323	0.308	0.307	0.319	0.313

,N2)	(8 ,8)			(8 ,27)			(8 ,64)			(8 ,200)			(27 ,27)		
IRE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
DF	0.434	0.102	0.268	0.412	0.058	0.235	0.466	0.040	0.253	0.448	0.038	0.243	0.446	0.048	0.247
DF	0.301	0.141	0.221	0.289	0.130	0.210	0.390	0.069	0.229	0.369	0.064	0.216	0.337	0.091	0.214
DF	0.202	0.178	0.190	0.118	0.231	0.174	0.105	0.253	0.179	0.088	0.279	0.184	0.157	0.172	0.164
DF	0.224	0.178	0.201	0.150	0.207	0.178	0.178	0.199	0.188	0.180	0.207	0.193	0.157	0.172	0.165
DF	0.212	0.184	0.198	0.146	0.198	0.172	0.143	0.199	0.171	0.137	0.194	0.165	0.157	0.175	0.166
DF	0.235	0.184	0.209	0.183	0.191	0.187	0.207	0.175	0.191	0.220	0.170	0.195	0.160	0.177	0.169

N2)	(27 ,64)			(27 ,200)			(64 ,64)			(64 ,200)			(200 ,200)		
IRE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
F	0.467	0.037	0.252	0.503	0.027	0.265	0.469	0.043	0.256	0.476	0.031	0.253	0.460	0.032	0.246
F	0.379	0.062	0.221	0.397	0.051	0.224	0.377	0.061	0.219	0.426	0.047	0.237	0.418	0.048	0.233
F	0.130	0.209	0.167	0.085	0.271	0.178	0.162	0.159	0.160	0.102	0.234	0.168	0.165	0.153	0.159
F	0.139	0.200	0.170	0.106	0.236	0.171	0.162	0.160	0.161	0.108	0.225	0.167	0.165	0.153	0.159
F	0.158	0.185	0.171	0.125	0.206	0.166	0.163	0.167	0.162	0.129	0.197	0.163	0.164	0.154	0.159
F	0.162	0.181	0.172	0.136	0.193	0.165	0.164	0.161	0.163	0.131	0.198	0.164	0.165	0.155	0.160

TABLE 5

RESULTS FOR THE INVERSE HYPERBOLIC SINE NORMAL DISTRIBUTION
TOTAL PROPORTION OF MISCLASSIFIED OBSERVATIONS

N2)	(8 ,8)			(8 ,27)			(8 ,64)			(8 ,200)			(27 ,27)		
RE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
F	0.500	0.262	0.381	0.430	0.253	0.341	0.487	0.214	0.351	0.463	0.217	0.340	0.448	0.220	0.334
F	0.542	0.336	0.439	0.495	0.317	0.406	0.580	0.262	0.421	0.558	0.245	0.401	0.528	0.243	0.386
F	0.382	0.328	0.355	0.321	0.340	0.330	0.325	0.322	0.324	0.301	0.359	0.330	0.335	0.298	0.316
F	0.450	0.341	0.396	0.348	0.350	0.349	0.343	0.336	0.339	0.350	0.350	0.350	0.345	0.296	0.320
F	0.376	0.329	0.353	0.339	0.323	0.331	0.342	0.305	0.323	0.330	0.321	0.326	0.335	0.301	0.318
F	0.457	0.334	0.396	0.372	0.338	0.355	0.398	0.308	0.353	0.407	0.304	0.356	0.342	0.302	0.322

N2)	(27 ,64)			(27 ,200)			(64 ,64)			(64 ,200)			(200 ,200)		
RE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
F	0.440	0.219	0.329	0.482	0.177	0.329	0.471	0.226	0.323	0.430	0.208	0.319	0.444	0.206	0.325
F	0.465	0.277	0.371	0.572	0.169	0.371	0.573	0.179	0.376	0.540	0.183	0.361	0.592	0.150	0.371
F	0.303	0.321	0.312	0.282	0.336	0.309	0.317	0.301	0.309	0.265	0.348	0.307	0.301	0.322	0.312
F	0.311	0.324	0.318	0.275	0.354	0.315	0.309	0.317	0.313	0.255	0.359	0.307	0.305	0.320	0.312
F	0.321	0.307	0.314	0.324	0.296	0.310	0.318	0.304	0.311	0.287	0.325	0.306	0.303	0.321	0.312
F	0.335	0.318	0.327	0.337	0.304	0.321	0.304	0.326	0.315	0.293	0.323	0.308	0.307	0.319	0.313

N2)	(8 ,8)			(8 ,27)			(8 ,64)			(8 ,200)			(27 ,27)		
RE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
F	0.413	0.126	0.269	0.370	0.076	0.223	0.429	0.054	0.242	0.402	0.053	0.227	0.409	0.063	0.236
F	0.323	0.192	0.257	0.332	0.135	0.234	0.431	0.091	0.261	0.401	0.082	0.242	0.369	0.115	0.242
F	0.202	0.178	0.190	0.118	0.231	0.174	0.105	0.253	0.179	0.088	0.279	0.184	0.157	0.172	0.164
F	0.224	0.178	0.201	0.150	0.207	0.178	0.178	0.199	0.188	0.180	0.207	0.193	0.157	0.172	0.165
F	0.212	0.184	0.198	0.146	0.198	0.172	0.143	0.199	0.171	0.137	0.194	0.165	0.157	0.175	0.166
F	0.235	0.184	0.209	0.183	0.191	0.187	0.207	0.175	0.191	0.220	0.170	0.195	0.160	0.177	0.169

N2)	(27 ,64)			(27 ,200)			(64 ,64)			(64 ,200)			(200 ,200)		
RE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
F	0.431	0.044	0.237	0.457	0.036	0.247	0.473	0.053	0.238	0.428	0.038	0.233	0.411	0.045	0.228
F	0.399	0.084	0.241	0.432	0.062	0.247	0.415	0.077	0.247	0.437	0.063	0.250	0.440	0.059	0.249
F	0.130	0.209	0.169	0.085	0.271	0.178	0.162	0.159	0.160	0.102	0.234	0.168	0.165	0.153	0.159
F	0.139	0.200	0.170	0.106	0.236	0.171	0.162	0.160	0.161	0.108	0.225	0.167	0.165	0.153	0.159
F	0.155	0.185	0.171	0.125	0.206	0.166	0.163	0.160	0.162	0.129	0.197	0.163	0.164	0.154	0.159
F	0.162	0.181	0.172	0.136	0.193	0.165	0.164	0.161	0.163	0.131	0.198	0.164	0.165	0.155	0.160

the ILDF and IQDF were approximately the same and smaller than that for the rank type procedures.

Summary

For these types of non-normal distributions, the LDF and QDF based upon the original data were clearly inappropriate as a classification procedure. The proportion of misclassified observations was substantially larger than the optimal rate, and also substantially larger than the overall error rate of the nonparametric procedures. The discrepency between $\hat{P}(1/2)$ and $\hat{P}(2/1)$ was substantial.

When u = 1, the procedures based on the ranks and on the normal scores were approximately equal. As μ increased (i.e. $\mu = 2$) and the distance between the two distributions increased, the procedures based on the inverse normal scores transformation classified the data more appropriately based on the criteria of $\hat{P}(1/2) = \hat{P}(2/1)$.

Logit Normal Distribution

Table 6 presents the results for the logit normal distribution. As outlined previously, the results for the rank and normal scores type procedures were identical for all of the distributions. Therefore, the only difference concerns whether the procedures based on the LDF and QDF for the original data were appropriate for the data. For the logit normal samples, the LDF and QDF classified the data equally as well as the four nonparametric procedures. In fact, the results for this distribution were almost identical to the results for the bivariate normal distribution. For that reason, a discussion of these results is omitted and the reader should consult the section outlining the bivariate normal results.

TABLE 6

RESULTS FOR THE LOGIT NORMAL DISTRIBUTION
TOTAL PROPORTION OF MISCLASSIFIED OBSERVATIONS

(N1,N2)		(8 ,8)			(8 ,27)			(8 ,64)			(8 ,200)			(27 ,27)		
MEAN	PROCEDURE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
1	LDF	0.367	0.338	0.353	0.319	0.336	0.328	0.326	0.315	0.320	0.315	0.340	0.328	0.297	0.336	0.316
	QDF	0.416	0.368	0.392	0.353	0.345	0.349	0.364	0.327	0.345	0.382	0.324	0.353	0.286	0.358	0.322
	RLDF	0.382	0.328	0.355	0.321	0.340	0.330	0.325	0.322	0.324	0.301	0.359	0.330	0.335	0.298	0.316
	RQDF	0.450	0.341	0.396	0.348	0.350	0.349	0.343	0.336	0.339	0.350	0.350	0.350	0.345	0.296	0.320
	ILDF	0.376	0.329	0.353	0.339	0.323	0.331	0.342	0.305	0.323	0.330	0.321	0.326	0.335	0.301	0.318
	IQDF	0.457	0.334	0.396	0.372	0.338	0.355	0.398	0.308	0.353	0.407	0.304	0.356	0.342	0.302	0.322

(N1,N2)		(27 ,64)			(27 ,200)			(64 ,64)			(64 ,200)			(200 ,200)		
MEAN	PROCEDURE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
1	LDF	0.286	0.341	0.313	0.295	0.323	0.309	0.279	0.341	0.310	0.269	0.342	0.306	0.275	0.350	0.313
	QDF	0.287	0.361	0.324	0.293	0.339	0.316	0.243	0.387	0.315	0.255	0.360	0.308	0.256	0.373	0.314
	RLDF	0.303	0.321	0.312	0.282	0.336	0.309	0.317	0.301	0.309	0.265	0.348	0.307	0.301	0.322	0.312
	RQDF	0.311	0.324	0.318	0.275	0.354	0.315	0.309	0.317	0.313	0.255	0.359	0.307	0.305	0.320	0.312
	ILDF	0.321	0.307	0.314	0.324	0.296	0.310	0.318	0.304	0.311	0.287	0.325	0.306	0.303	0.321	0.312
	IQDF	0.335	0.318	0.327	0.337	0.304	0.321	0.304	0.326	0.315	0.293	0.323	0.308	0.307	0.319	0.313

(N1,N2)		(8 ,8)			(8 ,27)			(8 ,64)			(8 ,200)			(27 ,27)		
MEAN	PROCEDURE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
2	LDF	0.125	0.232	0.178	0.117	0.235	0.176	0.118	0.227	0.173	0.109	0.235	0.172	0.110	0.232	0.171
	QDF	0.155	0.210	0.182	0.164	0.208	0.186	0.202	0.182	0.192	0.131	0.192	0.186	0.118	0.222	0.170
	RLDF	0.202	0.178	0.190	0.118	0.231	0.174	0.105	0.253	0.179	0.088	0.279	0.184	0.157	0.172	0.164
	RQDF	0.224	0.178	0.201	0.150	0.207	0.178	0.178	0.199	0.188	0.180	0.207	0.193	0.157	0.172	0.165
	ILDF	0.212	0.184	0.198	0.146	0.198	0.172	0.143	0.199	0.171	0.137	0.194	0.165	0.157	0.175	0.166
	IQDF	0.235	0.184	0.209	0.183	0.191	0.167	0.207	0.175	0.191	0.220	0.170	0.195	0.160	0.177	0.169

(N1,N2)		(27 ,64)			(27 ,200)			(64 ,64)			(64 ,200)			(200 ,200)		
MEAN	PROCEDURE	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P	P(1/2)	P(2/1)	P
2	LDF	0.117	0.230	0.174	0.101	0.235	0.168	0.104	0.233	0.169	0.100	0.237	0.169	0.101	0.231	0.166
	QDF	0.134	0.210	0.172	0.125	0.207	0.166	0.109	0.223	0.166	0.105	0.229	0.167	0.108	0.221	0.164
	RLDF	0.130	0.209	0.169	0.085	0.271	0.178	0.162	0.159	0.160	0.102	0.234	0.168	0.165	0.153	0.159
	RQDF	0.139	0.200	0.170	0.106	0.236	0.171	0.162	0.160	0.161	0.108	0.225	0.167	0.165	0.153	0.159
	ILDF	0.158	0.185	0.171	0.125	0.206	0.166	0.163	0.160	0.162	0.129	0.197	0.163	0.164	0.154	0.159
	IQDF	0.162	0.181	0.172	0.136	0.193	0.165	0.164	0.161	0.163	0.131	0.198	0.164	0.165	0.155	0.160

Conclusions

With samples drawn from bivariate normal distributions with equal covariance matrices, the proportion of observations misclassified using the LDF procedure based on the ranks or the normal scores of the data was not considerably different from that for the LDF based on the original data. Furthermore, the RLDF and IQDF proportions of misclassification were almost equivalent to the optimal values in the bivariate normal case. Hence, it is to be expected that they would also be approximately equal to the optimal value for the non-normal situations because those procedures are not affected by the transformation from normality.

For the non-normal distributions, the LDF and QDF were clearly inappropriate. They type of non-normality,however,appeared to have some effect on the performance of those procedures. The LDF and QDF suffered least when the distribution was bounded above and below (i.e. for the finite range logit normal distribution). When the range was semi-infinite or infinite, there was substantial increase in the overall error rate and the inflation/deflation was considerable.

A discussion of sample size is appropriate. With the normally distributed samples, little was gained by using sample sizes larger than 27 for any of the procedures. This was also true for the procedures based on the ranks and inverse normal scores for the non-normal data. This result contrasts with the nonparametric methods studied by Koffler & Penfield(1979) which required a fairly large sample size and showed improvements as the sample size increased beyond 64.

When $n_1 = n_2$, i.e. the sample sizes for estimating the density function parameters were equal, the four nonparametric procedures classified the data equally as well. When the sample sizes were unequal, the procedures based on the inverse normal scores tended to more effectively classify the data. For those situations the procedures based upon the ranks exhibited an inflation/deflation effect.

In summary, when the distributions are normal, the rank and inverse normal scores metods are effective substitutes for the LDF and QDF. When the populations are non-normal, the LDF methods based on the ranks or the inverse normal scores are more effective than the LDF or QDF methods based on the raw data. Finally, when the criterion sample sizes are unequal, the inverse normal scores approach is more desirable than the rank approach. When the criterion sample sizes are equal, either of the two procedures can be used.

REFERENCES

Anderson, T.W. Classification by multivariate analysis. Psychometrika, 1951, 16, 31-50.

Anderson, T.W. Some nonparametric multivariate procedures based on statistically equivalent blocks. Multivariate Analysis. Proceedings of the International Symposium, Dayton, Ohio, 1965. New York: Academic Press, 1966, 5-27.

Conover, W.J. Practical Nonparametric Statistics. New York: John Wiley, 1971.

Conover, W.J. & Iman, R.L. On some alternative procedures using ranks for the analysis of experimental designs. Communications in Statistics, Theory & Methods. 1976, 14, 1349-1368.

Fix, E. & Hodges J.L. Nonparametric discrimination: Consistency properties. Lackland Air Force Base, Texas: U.S. School of Aviation Medicine, 1951.

Gessaman, M.P. & Gessaman, P.H. A comparison of some multivariate discrimination procedures. Journal of the American Statistical Association, 1972, 67, 468-472.

Hajek, J. & Sidak, Z. Theory of Rank Tests. Prague: Academic Press, 1967.

Hoeffding, W. 'Optimum' nonparametric tests. Proceedings of the Second Berkeley Symposium on Mathematical Statistics & Probability. Berkeley & Los Angeles: University of California Press, 1951.

Hoel, P.G. & Peterson, R.P. A solution to the problem of optimum classification. Annals of Mathematical Statistics, 1949, 20, 433-438.

Iman, R.L. A power study of a rank transform for the two-way classification model when interaction may be present. The Canadian Journal of Statistics Section C: Applications, 1974, 2, 227-239.

Iman, R.L. & Conover, W.J. The use of the rank transform in regression. Technometrics, 1977.

Johnson, N.L. Systems of frequency curves generated by methods of translation. Biometrika, 1949, 36, 149-176.

Johnson, M.E. & Ramberg, J.S. Transformations of the multivariate normal distribution with applications to simulation. Technical report LA-UR-77-2295, Los Alamos Scientific Laboratory, New Mexico, 1977.

Koffler, S.L. & Penfield, D.A. Nonparametric discrimination procedures for non-normal distributions. Journal of Statistical Computation & Simulation. 1979, 8, 281-299.

Kossack, C.F. & Henschke, C.I. Introduction to Statistics and Computer Programming. San Francisco: Holden-Day, Inc. 1975.

Lachenbruch, P.A., Sneeringer, C. & Revo, L.T. Robustness of the linear and quadratic discriminant function to certain types of non-normality. Communications in Statistics, 1973, 1, 39-56.

Lehmann, E.L. Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day, Inc. 1975.

Loftsgaarden, D.O. & Quesenberry, C.P. A nonparametric estimate of a multivariate density function. Annals of Mathematical Statistics, 1965, 36, 1049-1051.

Marks, S. & Dunn, O.J. Discriminant functions when covariance matrices are unequal. Journal of the American Statistical Association. 1974, 69, 555-559.

Marascuilo, L.A. Large sample multiple comparisons. Psychological Bulletin. 1966, 65, 280-290.

McSweeney. M. & Penfield, D.A. The normal scores test for the c-sample problem. The British Journal of Mathematical and Statistical Psychology. 1969, 22, 177-192.

Pratt, J.W. Robustness of some procedures for the two-sample location problem. Journal of the American Statistical Association. 1964, 59, 665-680.

Puri, M.L. Asymptotic efficiency of a class of c-sample tests. Annals of Mathematical Statistics, 1964, 35, 102-121.

Ramberg, J.S. & Schmeiser, B.W. An approximate method for generating symmetric random variables. Communications of the Association for Computer Machinery. 1972, 15, 987-990.

Terry, M.E. Some rank order tests which are most powerful against specific parametric alternatives. Annals of Mathematical Statistics, 1952, 23, 346-366.

Van der Waerden, B.L. Order tests for the two-sample problem and their power. Indagationes Mathematicae, 1952, 14, 453-458.

Van der Waerden, B.L. Order tests for the two-sample problem. Indagationes Mathematicae, 1953, 15, 303-316.

Van der Waerden, B.L. The computation of the X-distribution. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. Berkeley & Los Angeles: University of California Press, 1956.

Welch, B.L. Notes on discriminant functions. Biometrika, 1939, 31, 218-220.